FCProfiler: Structured and Deterministic Pipeline for Recipe-Level Nutrient Estimation

Siyoon Kim
sykim2601@knou.ac.kr

Abstract

Food and nutrition question answering involves resolving ambiguous ingredient terminology and diverse household measurement expressions, and converting them into representations compatible with nutrient databases. In FoodBench-QA, recipe-level nutrient estimation requires consistent handling of heterogeneous and imprecise measurement descriptions. We propose FoodComponentProfiler (FCProfiler), a deterministic pipeline that treats nutrient estimation as a structured measurement resolution problem. The pipeline is composed of multiple stages, including parsing, normalization, unit canonicalization, gram conversion, and nutrient estimation, with each step designed to remain transparent and traceable. Unit canonicalization combines rule-based standards with data-driven unit expansion from large-scale recipe corpora, enabling broader coverage of real-world measurement variations. Gram conversion grounds quantities in ingredient-specific portion information, enabling accurate and traceable mass computation. Experimental results show that accurate nutrient estimation mainly depends on reliable unit normalization and ingredient-specific measurement conversion. Additionally, FCProfiler achieves performance comparable to FoodyLLM, demonstrating that explicit measurement grounding serves as an effective alternative to implicit reasoning. The proposed methodology preserves interpretability while maintaining strong performance in food and nutrition question answering.

Keywords: recipe-level nutrient estimation, measurement normalization, deterministic modeling

1. Introduction

Food and nutrition question answering involves challenges beyond general-purpose QA. It requires resolving ambiguous ingredient terminology, heterogeneous measurement expressions, and aligning textual descriptions with structured food composition databases (Min et al., 2019).

FoodBench-QA shared task, held as part of CL4Health at LREC 2026, addresses these challenges using nutrient databases and food ontologies. In FoodBench-QA, the problem is defined as a grounded food understanding task, where recipe-level nutrient estimation is evaluated under EU Regulation 1169/2011 tolerance standards (Bairati, 2017). This setting places strong emphasis on accurate mass normalization and database-grounded nutrient estimation.

In this context, consistent handling of measurement expressions becomes critical. Household units such as “cup,” “tablespoon,” or “pinch” are ingredient-dependent and loosely defined, and even small inconsistencies in unit interpretation or gram conversion can lead to substantial deviations in nutrient estimates, particularly for high-density nutrients.

While large language models (LLMs) demonstrate strong generative performance, end-to-end generation provides limited control over numerical normalization and intermediate mass computation, which are essential for reproducible nutrient estimation (Minaee et al., 2025).

To address these challenges, we develop FoodComponentProfiler (FCProfiler), a deterministic pipeline for recipe-level nutrient estimation. The approach decomposes the task into explicit stages: unit canonicalization following standardized measurement rules such as NIST SP 811 (Thompson and Taylor, 2008), ingredient-specific gram conversion grounded in USDA FoodData Central (U.S. Department of Agriculture, Agricultural Research Service, Beltsville Human Nutrition Research Center, 2026), and structured nutrient estimation. To improve coverage of informal and diverse measurement expressions, unit canonicalization is further expanded using corpus-driven mining from Recipe1M+ (Marin et al., 2021).

This structured design leads to the following key contributions:

- We design a deterministic and interpretable pipeline for recipe-level nutrient estimation, explicitly separating unit canonicalization, gram conversion, and nutrient estimation.
- We demonstrate through ablation studies that reliable unit canonicalization and ingredient-specific gram conversion are the primary factors driving nutrient estimation accuracy on FoodBench-QA dataset (Eftimov et al., 2025b).
- We show that the proposed pipeline attains performance competitive with FoodyLLM on FoodBench dataset through explicit measurement grounding (Eftimov et al., 2025a).

Proceedings of the Third Workshop on Patient-Oriented Language Processing (CL4Health), pages 369–374

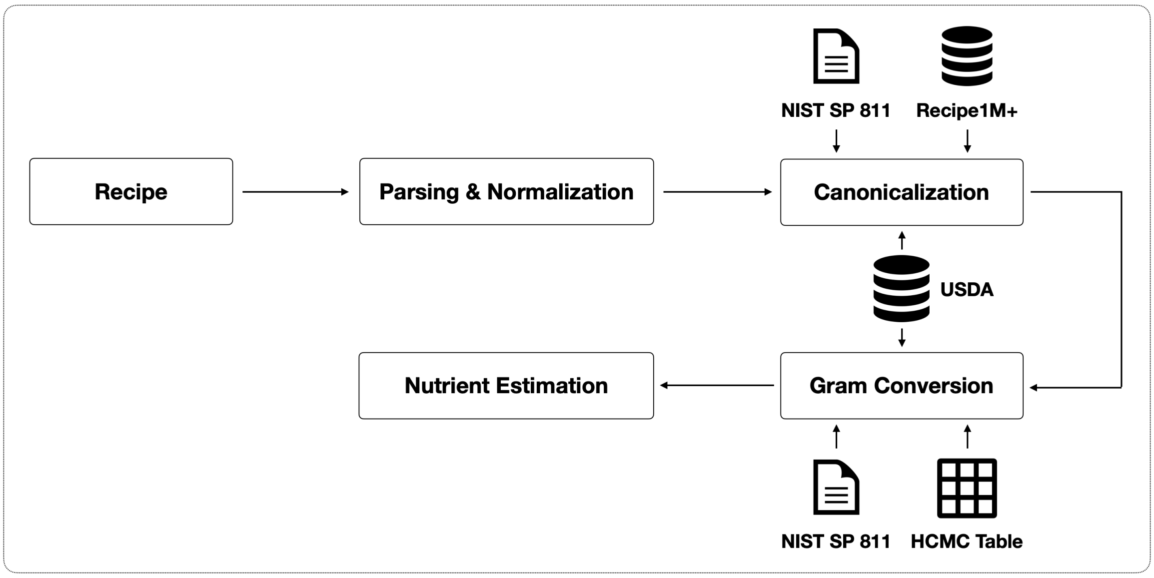

Figure 1: FCProfiler pipeline overview

2. Related Work

Prior work on food-related nutrient estimation has been explored through various paradigms, including image-based and text-based approaches, as well as recent LLM-based and multimodal approaches.

Image-based approaches to nutrient estimation involve food image segmentation, portion size estimation, volume-to-mass conversion, and calorie estimation (Pouladzadeh et al., 2014; Myers et al., 2015). Nutrient values can be derived through database lookup based on food type and mass. However, these approaches primarily focus on calorie estimation and rely on intermediate physical approximations, making them sensitive to error propagation and limiting their applicability to accurate nutrient estimation.

Text-based approaches have explored predicting nutrient values directly from textual descriptions. P-NUT demonstrated that macronutrient content can be predicted from short food descriptions using text embeddings and clustering-based regression models (Ispirova et al., 2020). However, this approach relies on limited textual descriptions and does not incorporate ingredient-level structure, quantity information, or explicit grounding in food composition databases, reducing its applicability to realistic recipe scenarios. To address these limitations, subsequent studies extended the problem to structured recipe data (Ispirova et al., 2022), extracting ingredients, quantities, and measurement units using rule-based and dictionary-based methods and mapping them to food composition databases such as USDA FoodData Central. While this enables ingredient-aware representations and incorporation of measurement information, nutrient estimation is still treated as a learned prediction task rather than an explicit measurement resolution process.

Recent approaches have adopted LLMs and multimodal approaches to address multiple food-related tasks in a unified framework. FoodyLLM integrates nutrient estimation, traffic-light classification, and food entity linking as a multi-task instruction learning problem (Gjorgjevikj et al., 2026), while recent multimodal models further incorporate visual and textual inputs for holistic food understanding (Yin et al., 2025). However, these approaches rely on implicit reasoning within end-to-end architectures and do not provide explicit control over intermediate steps such as unit normalization and ingredient-specific gram conversion, which are critical for reproducible and numerically consistent nutrient estimation.

In contrast to these approaches, FCProfiler formulates recipe-level nutrient estimation as a structured measurement resolution framework. Rather than predicting nutrient values directly, the proposed pipeline explicitly models ingredient parsing, unit canonicalization, gram conversion, and database-grounded computation, enabling interpretable, reproducible, and numerically traceable nutrient estimation.

3. Methodology

FCProfiler is decomposed into structured stages: ingredient parsing and normalization, unit canonicalization, gram conversion, and nutrient estimation. Each stage produces explicit intermediate representations, enabling traceability from raw ingredient text to final nutrient values.

Dataset	Split	Protein	Sugars	Fat	Saturates
Ingredients Only	Train	0.9511	0.9065	0.9146	0.9291
Ingredients Only	Test	0.9461	0.9200	0.9224	0.9336
Title + Ingredients	Train	0.9507	0.9070	0.9142	0.9294
Title + Ingredients	Test	0.9461	0.9178	0.9236	0.9325

Table 1: Nutrient estimation accuracy of FCProfiler on FoodBench-QA train and test datasets under tolerance-based evaluation protocol

Dataset	Model	Protein	Sugars	Fat	Saturates
Ingredients Only	FoodyLLM	$\mathbf{0.972 \pm 0.002}$	0.917 ± 0.005	0.920 ± 0.004	$\mathbf{0.938 \pm 0.004}$
Ingredients Only	FCProfiler	0.945 ± 0.002	$\mathbf{0.920 \pm 0.003}$	$\mathbf{0.922 \pm 0.002}$	0.934 ± 0.001
Title + Ingredients	FoodyLLM	$\mathbf{0.967 \pm 0.002}$	0.915 ± 0.008	0.912 ± 0.003	$\mathbf{0.934 \pm 0.004}$
Title + Ingredients	FCProfiler	0.946 ± 0.002	$\mathbf{0.919 \pm 0.003}$	$\mathbf{0.924 \pm 0.002}$	0.933 ± 0.002

Table 2: Comparative evaluation of FCProfiler and FoodyLLM on FoodBench test dataset under tolerance-based evaluation protocol

3.1. Ingredient Parsing & Normalization

Given an input recipe, FCProfiler extracts structured tuples consisting of an ingredient name, quantity, and measurement unit. Ingredient parsing is performed using a rule-based approach based on regular expressions and token-level segmentation. The extracted ingredient names are then normalized through string-level preprocessing to produce representations for subsequent stages.

3.2. Unit Canonicalization

Household measurement expressions exhibit substantial variability (e.g., “tbsp”, “T”, “fl. oz”, “pinch”). To standardize these forms, FCProfiler applies layered canonicalization combining standardized measurement conventions (e.g., NIST SP 811), unit aliases derived from USDA FoodData Central, and corpus-driven unit expansion from Recipe1M+. This process maps heterogeneous expressions to consistent unit representations compatible with nutrient databases.

3.3. Gram Conversion

Ingredient quantities are converted to grams using a prioritized conversion strategy. Direct SI units are deterministically mapped. For volume- and count-based units, conversion is performed using ingredient-specific portion information from USDA FoodData Central or standardized mappings from Household Cooking Measurements Conversion (HCMC) table, ensuring ingredient-specific mass estimation rather than relying on global density assumptions.

3.4. Nutrient Estimation

After gram conversion, nutrient values are retrieved from USDA FoodData Central and aggregated at the recipe level. Final values are normalized per 100g following the shared task protocol. All computations are deterministic and database-grounded.

4. Experimental Results

4.1. Datasets

Experiments are conducted on FoodBench-QA dataset, which consists of English recipe text, evaluated under EU Regulation 1169/2011 tolerance standards. FoodBench-QA dataset provides two derived datasets for nutrient estimation task: *Ingredients Only* and *Title + Ingredients*. Each dataset includes predefined train and test splits, and evaluation is performed using tolerance-based nutrient accuracy for protein, sugars, fat, and saturates.

For comparison with FoodyLLM, we additionally evaluate on FoodBench dataset, which follows the same subset construction with *Ingredients Only* and *Title + Ingredients* and the same tolerance-based evaluation protocol for all nutrients. The proposed pipeline is evaluated using only the provided test splits from five independently sampled datasets.

Standardized measurement guidelines (NIST SP 811) and food-related resources, including USDA FoodData Central, Recipe1M+, and HCMC, are employed throughout unit canonicalization and gram conversion.

4.2. Results

Table 1 represents the final tolerance-based nutrient accuracy of FCProfiler under both datasets: *Ingredients Only* and *Title + Ingredients*. Across both train and test splits, FCProfiler achieves consistently high accuracy for all nutrients. On the test split, the full configuration with canonicalization and layered gram conversion reaches 0.9461 for pro-

Unit Canonicalization Strategy	Split	Protein	Sugars	Fat	Saturates
None	Train	0.5378	0.4259	0.3251	0.3935
NIST SP	Train	0.9332	0.8843	0.8879	0.9088
NIST SP + USDA	Train	0.9441	0.8981	0.9032	0.9201
NIST SP + USDA + Recipe1M+	Train	**0.9511**	**0.9065**	**0.9146**	**0.9291**
None	Test	0.5039	0.4045	0.2893	0.3538
NIST SP	Test	0.9404	0.9134	0.9146	0.9287
NIST SP + USDA	Test	0.9431	0.9152	0.9170	0.9295
NIST SP + USDA + Recipe1M+	Test	**0.9461**	**0.9200**	**0.9224**	**0.9336**

Table 3: Ablation results of unit canonicalization strategies for nutrient estimation

Gram Conversion Strategy	Split	Protein	Sugars	Fat	Saturates
Direct conversion only	Train	0.6282	0.5128	0.4252	0.5088
+ USDA portion-based conversion	Train	0.9320	0.8901	0.8959	0.9140
+ HCMC table-based conversion	Train	0.9441	0.8981	0.9032	0.9201
+ Heuristic fallback	Train	**0.9511**	**0.9065**	**0.9146**	**0.9291**
Direct conversion only	Test	0.5795	0.4717	0.3672	0.4550
+ USDA portion-based conversion	Test	0.9285	0.9036	0.9073	0.9227
+ HCMC table-based conversion	Test	0.9431	0.9152	0.9170	0.9295
+ Heuristic fallback	Test	**0.9461**	**0.9200**	**0.9224**	**0.9336**

Table 4: Ablation results of gram conversion strategies for nutrient estimation

tein, 0.9200 for sugars, 0.9224 for fat, and 0.9336 for saturates.

4.3. Comparison with FoodyLLM

Table 2 compares FCProfiler with FoodyLLM on FoodBench test dataset. FCProfiler achieves comparable performance across all nutrients for both *Ingredients Only* and *Title + Ingredients* datasets. In particular, FCProfiler shows slightly higher accuracy on sugars and fat across both datasets. The reported results are expressed as mean and standard deviation over five independently sampled datasets, and FoodyLLM results are taken directly from the original paper. These results indicate that explicit measurement grounding can provide a competitive alternative to end-to-end neural reasoning for nutrient estimation.

4.4. Ablation Study

Tables 3 and 4 represent ablation results on *Ingredients Only* dataset of FoodBench-QA dataset.

4.4.1. Unit Canonicalization

When unit canonicalization is completely disabled, performance collapses across all nutrients, particularly on the test split. This confirms that resolving heterogeneous household measurement expressions is essential for meaningful nutrient estimation.

Applying NIST SP-based canonicalization alone produces a substantial improvement, with test accuracy above 0.91 for all nutrients. Incorporating USDA alias expansion and corpus-driven Recipe1M+ unit expansion yields further improvements, with the full configuration achieving the best results across all nutrients.

The most significant improvements are observed when transitioning from no canonicalization to NIST SP-based canonicalization, indicating that standardizing measurement representations is a key factor for tolerance-aware accuracy.

4.4.2. Gram Conversion

Using direct unit-to-gram conversion alone substantially reduces performance on the test split, with protein accuracy dropping to 0.5795 and fat to 0.3672. This indicates that direct conversion alone is insufficient for volume- and count-based units commonly found in recipes.

Incorporating USDA portion-based conversion significantly improves performance, raising test accuracy to 0.9285 for protein and 0.9073 for fat. Adding HCMC table-based conversion further increases protein accuracy to 0.9431 and fat to 0.9170. Finally, heuristic fallback mechanisms yield the best results, achieving 0.9461 for protein and 0.9224 for fat on the test split.

The combined strategy achieves the highest accuracy across all nutrients, indicating that nutrient estimation accuracy largely depends on accurate mass grounding.

5. Discussion

Accurate nutrient estimation requires resolving ingredient quantities into standardized gram values through ingredient-specific measurement interpretation. For instance, in "1 teaspoon wheat flour," the unit must be interpreted and converted into grams based on ingredient-specific portion information. In FCProfiler, this is resolved by mapping the unit to *tsp* and grounding it in USDA portion data, yielding 2.60g. This example underscores the importance of accurate unit canonicalization and ingredient-specific gram conversion for reliable nutrient estimation.

This observation extends consistently across ingredients. The same unit, such as teaspoon or cup, corresponds to different masses depending on the ingredient type, making fixed density assumptions inherently unreliable. By grounding conversions in ingredient-specific portion metadata, FCProfiler reduces systematic errors introduced by generic conversion rules.

Unit canonicalization is also essential as many recipe measurements are expressed in heterogeneous and non-standardized forms. The results in Table 3 show that NIST SP-based canonicalization alone leads to substantial performance improvements, indicating that mapping diverse measurement expressions into a consistent unit system is important for reliable gram conversion. Additional alias expansion and corpus-driven expansion provide incremental improvements, with their contribution expected to be more pronounced in more heterogeneous recipe corpora.

6. Conclusion

We propose FCProfiler, a deterministic and interpretable pipeline for recipe-level nutrient estimation. Rather than relying on generative models, the pipeline performs structured measurement processing and database-grounded computation. Experimental results indicate that unit canonicalization and ingredient-specific gram conversion are critical to nutrient estimation accuracy. In particular, portion-based grounding using USDA FoodData Central enables realistic volume-to-mass conversion without relying on global density assumptions, while preserving traceability throughout the pipeline. FCProfiler also achieves comparable performance to FoodyLLM on FoodBench dataset, indicating the effectiveness of explicit measurement grounding.

These findings indicate that reliable food and nutrition question answering depends on accurate unit canonicalization, ingredient-specific gram conversion, and database grounding. The source code is publicly available at `https://github.com/rbls-lab/FCProfiler` to support reproducibility, and includes implementation details and experimental configurations.

7. Limitations

Despite strong performance, portion ambiguity remains a practical limitation. Expressions such as *"1 pack noodles"* or *"1 serving sauce"* often lack standardized gram mappings. Although heuristic fallback strategies provide approximate mass values, these approximations introduce uncertainty in nutrient computation.

In the future, we would like to apply food category–aware constraints derived from nutrient distributions within USDA FoodData Central. By modeling expected nutrient ranges per food category, it may be possible to detect and adjust implausible computations caused by portion-related mass conversion errors while maintaining interpretability.

The same deterministic pipeline of FCProfiler can be extended to other FoodBench-QA shared tasks, including FSA traffic-light prediction and ontology-grounded food entity linking. This indicates that structured measurement resolution and database grounding provide a generalizable basis for food and nutrition question answering.

8. Acknowledgements

We are grateful to Dongwoo Kim from Korea National Open University for his insightful feedback, and to Kyunghee An of Seongseo Public Health Subcenter and Sunok Kim of Soonok Bakery for their helpful comments and practical insights. We also thank the anonymous reviewers for their valuable comments.

9. Ethics Statement

FCProfiler is evaluated on publicly available recipe data and uses open nutritional resources. The study does not involve personal or sensitive information. Nutrient estimates are database-grounded approximations and should not be interpreted as medical advice.

10. Bibliographical References

Lorenzo Bairati. 2017. The food consumer's right to information on product country of origin: trends and outlook, beyond eu regulation 1169/2011. *Journal of European Consumer and Market Law*, 6(1):9–16.

Ana Gjorgjevikj, Matej Martinc, Gjorgjina Cenikj, Riste Stojanov, Jan Drole, Gordana Ispirova, Giulia Menichetti, Nives Ogrinc, Dimitar Trajanov, Sašo Džeroski, Barbara Koroušić Seljak, and Tome Eftimov. 2026. Large language models in food and nutrition science: Opportunities, challenges, and the case of foodyllm. *Current Research in Food Science*, 12:101351.

Gordana Ispirova, Tome Eftimov, and Barbara Koroušić Seljak. 2020. P-nut: Predicting nutrient content from short text descriptions. *Mathematics*, 8(10).

Gordana Ispirova, Tome Eftimov, and Barbara Korousic Seljak. 2022. Predefined domain specific embeddings of food concepts and recipes: A case study on heterogeneous recipe datasets . In *2022 IEEE International Conference on Big Data (Big Data)*, pages 4074–4083.

Weiqing Min, Shuqiang Jiang, Linhu Liu, Yong Rui, and Ramesh Jain. 2019. A survey on food computing. *ACM Computing Surveys*, 52(5).

Shervin Minaee, Tomas Mikolov, Narjes Nikzad, Meysam Chenaghlu, Richard Socher, Xavier Amatriain, and Jianfeng Gao. 2025. Large language models: A survey. ArXiv preprint arXiv:2402.06196.

Austin Myers, Nick Johnston, Vivek Rathod, Anoop Korattikara, Alex Gorban, Nathan Silberman, Sergio Guadarrama, George Papandreou, Jonathan Huang, and Kevin Murphy. 2015. Im2calories: towards an automated mobile vision food diary. In *ICCV*.

Parisa Pouladzadeh, Shervin Shirmohammadi, and Rana Al-Maghrabi. 2014. Measuring calorie and nutrition from food image. *IEEE Transactions on Instrumentation and Measurement*, 63(8):1947–1956.

Ambler Thompson and Barry Taylor. 2008. Guide for the use of the international system of units (si). Special Publication (NIST SP), National Institute of Standards and Technology.

Yuehao Yin, Huiyan Qi, Bin Zhu, Jingjing Chen, Yu-Gang Jiang, and Chong-Wah Ngo. 2025. Foodlmm: A versatile food assistant using large multi-modal model. *IEEE Transactions on Multimedia*, 27:6949–6961.

11. Language Resource References

Eftimov, Tome and Gjorgjevikj, Ana and Martinc, Matej and Cenikj, Gjorgjina and Darko, Sasanski and Stojanov, Riste and Džeroski, Sašo and Koroušić Seljak, Barbara. 2025a. *FoodBench: A Domain-Targeted Question-Answering Dataset for Food and Nutrition Applications*. Zenodo.

Eftimov, Tome and Gjorgjevikj, Ana and Martinc, Matej and Cenikj, Gjorgjina and Džeroski, Sašo and Koroušić Seljak, Barbara. 2025b. *FoodBench-QA: Shared Task on Grounded Food & Nutrition Question Answering*. Accessed: 2026-02-18.

Marin, Javier and Biswas, Aritro and Ofli, Ferda and Hynes, Nicholas and Salvador, Amaia and Aytar, Yusuf and Weber, Ingmar and Torralba, Antonio. 2021. *Recipe1M+: A Dataset for Learning Cross-Modal Embeddings for Cooking Recipes and Food Images*.

U.S. Department of Agriculture, Agricultural Research Service, Beltsville Human Nutrition Research Center. 2026. *FoodData Central*. Accessed: 2026-02-18.